BULLETIN *of the*
Malaysian Mathematical
Sciences Society
http://math.usm.my/bulletin

Bull. Malays. Math. Sci. Soc. (2) **32***(1) (2009), 75–100*

Bayesian Approach for Joint Longitudinal and Time-to-Event Data with Survival Fraction

[1]Mohd Rizam Abu Bakar, [2]Khalid A. Salah,
[3]Noor Akma Ibrahim and [4]Kassim Haron

[1,4] Department of Mathematics, Universiti Putra Malaysia, Malaysia
[2]Department of Mathematics, Alquds University, Palestine
[3] Institute for Mathematical Research, Universiti Putra Malaysia, Malaysia
[1]mrizam@putra.upm.edu.my, [2]ksalah@science.alquds.edu,
[3]nakma@science.upm.edu.my, [4]kassim@science.upm.edu.my

Abstract. Many medical investigations generate both repeatedly-measured (longitudinal) biomarker and survival data. One of complex issue arises when investigating the association between longitudinal and time-to-event data when there are cured patients in the population, which leads to a plateau in the survival function $S(t)$ after sufficient follow-up. Thus, usual Cox proportional hazard model [11] is not applicable since the proportional hazard assumption is violated. An alternative is to consider survival models incorporating a cure fraction. In this paper, we present a new class of joint model for univariate longitudinal and survival data in presence of cure fraction. For the longitudinal model, a stochastic Integrated Ornstein-Uhlenbeck process will present, and for the survival model a semiparametric survival function will be considered which accommodate both zero and non-zero cure fractions of the dynamic disease progression. Moreover, we consider a Bayesian approach which is motivated by the complexity of the model. Posterior and prior specification needs to accommodate parameter constraints due to the non-negativity of the survival function. A simulation study is presented to evaluate the performance of the proposed joint model.

2000 Mathematics Subject Classification: 62F15, 62G99, 62M05, 62N01

Key words and phrases: Survival model, longitudinal model, cure rate model, fixed effects, random effects, Bayesian approach, integrated Ornstein-Uhlenbeck.

1. Introduction

Joint models for longitudinal and survival data have recently become quite popular in cancer, AIDS, and environmental health studies where a longitudinal biologic marker of the health-related outcome such as CD4 counts in HIV trials, immune response to vaccine, or quality of life in clinical trial can be an important predictor of survival or some other time-to-event. Often the observed longitudinal data are

Received: December 14, 2006; *Revised:* January 4, 2008.

incomplete or may be subject to error. Since such longitudinal markers (covariates) are measured with error, the analysis become more complex than one that treats these as fixed covariates in a survival model.

In many clinical studies, especially in cancer research, there are settings in which it is meaningful to consider the existence of a fraction of individuals who have little to no risk "cured" of experiencing the event of interest. For such failure time-data, a proportion of subjects are susceptible to, and others are not susceptible to, the target event. Empirical evidence to confirm this feature of the population is a long and stable plateau with heavy censoring at the tail of the Kaplan-Meire survival curve. With long term survivors, the usual Cox proportional hazard model [11] is not applicable since the proportional hazard assumption is violated. An alternative is to consider survival models incorporating a cure fraction, which, often referred to as a cure rate models.

The most popular type of cure rate model is the mixture model discussed by Berkson and Gage [1]. In this model, they assume a certain fraction θ of the population is "cured" and the remaining $(1-\theta)$ are not cured. The survival function for the entire population, denoted by $S(t)$ for this model is given by

$$S(t) = \theta + (1-\theta)S_1(t), \tag{1.1}$$

where $S_1(t)$ denotes the survivor function for the non-cured group in the population. Clearly (1.1) is improper since $S(\infty) = \theta$, and when covariates are included we have a different θ_i for each subject $i = 1, ...n$. A logistic regression structure for θ_i is usually given by

$$\theta_i = \frac{\exp\left(\delta^T Z_i\right)}{1 + \exp\left(\delta^T Z_i\right)}\ , \tag{1.2}$$

as assumed by Kuk and Chen [26], where θ_i is a probability and cannot be zero and Z_i is a vector of covariates. The standard mixture cure model has been extensively studied in the literature [2, 19, 29, 30, 34, 35, 41, 42] among others.

The main drawback of model (1.1) is that it lacks a proportional hazard structure if the covariates are modelled through θ, which is desirable property in carrying out covariates analysis. An alternative cure rate model, with a proportional hazards structure for the population, sometimes called the promotion time cure rate model discussed by Yakovlev and Tsodikov [40] and Chen *et al.* [7]. They developed their model by hypothesized the genesis of a cancer, which initiated by the mutation of some carcinogenic cells (metastasis-competent tumor cells). They assume a Poisson distribution to the number of carcinogenic cells left active after the initial treatment, denoted by N. Given N, the survival function for the entire population is given by

$$S(y) = \exp\left(-\lambda F(y)\right), \tag{1.3}$$

where λ is the mean of the Poisson count. The cure fraction is then given by

$$S(\infty) \equiv P(N = 0) = \exp(-\lambda).$$

Hence, model (1.3) can be written as

$$S(y) = \exp\left(-\lambda F(y)\right) = \exp\left\{\ln(\theta)F(y)\right\}, \tag{1.4}$$

where $F(t)$ is a proper cumulative distribution function represents the promotion time, that is, time to development of a detectable tumor mass. Common parametric choices for $F(t)$ are exponential [18] and Weibull distribution [8, 14]. Nonparametric choices have also been considered [22, 26, 34, 35]. There are also formulations of non-mixture cure models to incorporate long-term survivors [3, 4, 6, 7, 22, 39, 40].

A joint model is comprised of two linked submodels, one for the true longitudinal process and one for the failure time, along with additional specifications and assumptions that allow ultimately a full representation of the joint distribution of the observed data. In the statistical literature, maximum likelihood and Bayesian approaches have been used to obtain the estimates of the unknown joint model parameters. Law *et al.* [27] proposed a joint longitudinal and survival cure mixture model, they obtained maximum likelihood estimations of the parameters using Monte Carlo Expectation Maximization (MCEM) algorithm. However, this approach seems very difficult to apply to joint model. It would involve integrating the two component models over the distribution of the longitudinal process to obtain the marginal likelihood of the observed data, this requires numerical integration in a very high dimension space. This seems algebraically intractable. On the other hand, Ibrahim *et al.* [22], Brown and Ibrahim [3], Chen *et al.* [6], Chi and Ibrahim [8] and Cowling *et al.* [9] obtained the parameter estimation using the Bayesian approach. The Bayesian approach avoids the troubles in maximizing the likelihood function, making inferences based on the posterior density of the parameters. Hence, we will use this approach in our modelling, focusing on the estimation of the joint posterior density of all unknown model parameters.

Most joint models developed so far in the statistical literature have focused on time-to-event data models with no survival fraction, this motivate us to develop a survival model capable of accommodating a possible cure fraction survival function as well as linking relevant longitudinal markers to such a model. Then, the objective of this research is to develop a flexible longitudinal and survival joint cure rate model with a biological and medical meaning that it can be accommodate both zero and nonzero cure fraction and utilizing the latent class regression framework developed for single events, that allowed for event-specific survival processes. This can be done be achieving the following:

(1) To present a stochastic model for longitudinal measurements that has a capability of updating and predicting the longitudinal measurements at any time.
(2) To modify a semiparametric survival function that can control the parametricity at the tail of the survival curve as well as the nonparametrecity at the beginning and at the middle of the survival curve.
(3) To introduce some auxiliary variables (latent) that can acknowledge the convergence of sampling methods.

The development of the proposed joint model was primarily motivated by a clinical trial conducted by international cancer study group. A primary study goal is to investigate covariates (longitudinal and baseline) in terms of their effect on the probability of tumor cure and the progression time. This data set has been analyzed with cure models motivated by medical findings which suggest the existence of a

cured proportion. Failing to account for cure may lead to incorrect inferences thus motivating our main research.

The presentation of our proposed joint model in this article proceeds as follows. In Subsection 2.1, a longitudinal model is presented. In this model the response measurements are consisting of a combination of fixed effect, random effect, an Integrated Ornstein-Uhlenbeck (IOU) stochastic process and measurement error. In Subsection 2.2, we propose a cure rate model with *semiparametric* link function for the promotion time. Then these two models are combined to obtain a joint model. In Section 3, the likelihood function will be derived after introducing a *semiparametric* function. In Sections 4 and 5, we review model selection and simulation study. Then we conclude with discussion.

2. A new class of longitudinal and survival joint model

Given that the subject i, $i = 1, ..., n$, is observed at time t, that is $i \in \Re(t)$, where $\Re(t)$ is the risk set at time t. let Y_i and C_i denote the event time and censoring time respectively; let Z_i be a q-dimensional vector of baseline covariates and let $X_i(t)$ be the longitudinal process at time $t \geq 0$. Components of Z_i might also be time dependent covariates whose values are known exactly and that are "external" in the sense described by Kalbfleisch and Prentice [25]. Rather than observe V_i for all i, we observe only $V_i = \min(Y_i, C_i)$ and the censored indicator $\Delta_i = I(Y_i \leq C_i)$, which equals one for time-to-event and zero otherwise. Values of $X_i(t)$ are measured intermittently at times $t_{ij} \leq V_i$, $j = 1, ..., n_i$, for subject i, which may be different for each i; often; target values for the observations times are specified by a study protocol, although deviations from protocol are common. The observed longitudinal data on subject i may be subject to "error", thus we observed only $X_i^* = \{X_i^*(t_{i1}), ..., X_i^*(t_{in_i})\}^T$, whose elements may not exactly equal the corresponding $X_i(t_{ij})$.

A joint model is comprised of two linked submodels, one for the "true" longitudinal process $X_i^*(t_{ij})$ and one for the failure time Y_i, along with additional specifications and assumptions that allow ultimately a full representation of the joint distribution of the observed data $D_i = \{V_i, \Delta_i, X_i^*, t_i\}$, where $t_i = (t_{i1}, ...t_{in_i})^T$. The D_i's are taken to be independent across i, reflecting the belief that the disease process evolves independently for each subject. In the framework of joint modelling, we specifically assume that the time-to-event Y and vector of repeated measurements X, are conditionally independent given X^*.

2.1. The longitudinal process

In this article, following Taylor *et al.* [36], we consider the longitudinal process consisting of a combination of fixed effect, random effect, an Integrated Ornstein-Uhlenbeck (IOU) stochastic process and measurement error. In general, we assume that

$$\begin{cases} X_i(t_{ij}) = X_i^*(t_{ij}) + \epsilon_i(t_{ij}) \\ X_i^*(t_{ij}) = \varpi_1 U_1(t_{ij}) + \varpi_2 U_2(t_{ij}) + W_i(t_{ij}), \end{cases} \tag{2.1}$$

where $X_i(t_{ij})$ and $X_i^*(t_{ij})$ denote the observed and true value of a continuous time-dependent covariates (or disease marker) for subject i at time t_{ij}, $U_1(t_{ij})$ and

$U_2(t_{ij})$ represent fixed and random effects with respectively coefficients ϖ_1 and ϖ_2, $\epsilon_i(t_{ij})$ is measurement error and $W_i(t_{ij})$ are independent IOU stochastic process with covariance structure given by

$$\text{Cov}(W_i(t), W_i(s)) = \frac{\sigma^2}{2\alpha^3}\left(2\alpha \min(s,t) + e^{-\alpha s} + e^{-\alpha t} - 1 - e^{-\alpha|t-s|}\right); \tag{2.2}$$

where α and σ^2 are parameters. An appealing feature of model (2.1) is that it corresponds to a random effects model as α approaches zero and $\sigma^2/2\alpha$ maintains a constant. This can be seen directly from the observation under this circumstance, the IOU process is no more than a random effects model. Also, it is interesting to note that scaled Brownian motion is a special case of $W(t)$ in which α is infinitely large and $\sigma^2/2\alpha$ is constant. In general, this model is more flexible and plausible than a random effects model since it allows the marker to vary around a straight line and allows the data to determine the degree of this variation.

Note that $\text{Cov}(W_i(t), W_i(s))$ in (2.2) depends on s and t and not just on their difference, which can be described as a disadvantage of the IOU process, that is not a stationary, and hence it is necessary to have a natural time zero for each subject. In some applications it may be that there is no natural time zero or that time zero is not exactly known. Thus, following Taylor [35], we can overcoming this problem by analyzing the differences as follows:

Let Y_{iF_i} be the first measurements on subject i at time F_i, and let $D_{it} = Y_{it} - Y_{iF_i}$; for $t > F_i$. Then

$$D_{it} = b(t - F_i) + \beta(X_{it} - X_{iF_i}) + W_{it} - W_{iF_i} + \epsilon_{it} - \epsilon_{iF_i}, \tag{2.3}$$

and

$$\text{Cov}(D_{it_1}, D_{it_2}) = A + B + C, \tag{2.4}$$

where

$$\begin{aligned}
A &= (t_1 - F_i)(t_2 - F_i)\,\text{var}(b) - (X_{it_1} - X_{iF_i})(X_{it_2} - X_{iF_i})\,\text{var}(\beta)\\
B &= \text{Cov}(W_{it_1} - W_{iF_i}\ , W_{it_2} - W_{iF_i})\\
&= \frac{\sigma^2}{2\alpha^3}\left(2\alpha(t_1 - F_i) - 1 - e^{-\alpha(t_2 - t_1)} + e^{-\alpha(t_2 - F_i)} + e^{-\alpha(t_1 - F_i)}\right)\\
C &= \sigma_e^2\,(1 + I(t_1 = t_2)) = \begin{cases} 2\sigma_e^2 & \text{if} \quad t_1 = t_2 \\ \sigma_e^2 & \text{if} \quad t_1 \neq t_2\ , \end{cases}
\end{aligned}$$

for $t_1 \leq t_2$. By this assumption, we note that $\text{Cov}(D_{it_1}, D_{it_2})$ and hence $\text{Cov}(W_i(t), W_i(s))$, depends only on the difference in times, so it avoids the need to define natural time zero.

2.2. The time-to-event model

Motivated by the promotion time model, discussed by Yakovlev and Tsodikov [40], and following Chi and Ibrahim [8], we present a model which allows for a zero as well as a nonzero cure fraction. We propose such model by specifying an alternative mechanism for the characteristics of tumor growth. Instead of assuming the carcinogenic cells turn active only at the beginning of the study, we allow the possibility that active carcinogenic cells may occur at anytime after the start of the study. So

that, in addition to some carcinogenic cells remaining active after initial treatment, new carcinogenic cells are assumed to occur over time after this treatment. Thus the number of carcinogenic cells changes over time, and the risk of developing a cancer relapse becomes dynamic over time, the development of any active carcinogenic cells to become a detectable tumor then leads to relapse. Figure 1 shows a simple diagram to illustrate this idea. In the diagram, eight carcinogenic cells occurred during ten-years follow-up, and the patient relapses before the ninth year when the second metastasis-competent tumor cell first develops to become a real tumor. In terms of the statistical modelling, the promotion times for carcinogenic cells to become detectable tumor are assumed to be independent and identically distributed with a common distribution function, moreover, we consider a *semiparametric* version of the parametric cure rate model in (1.4). A stochastic nonhomogeneous Poisson process is also introduced to model the variation of the number of carcinogenic cells over time.

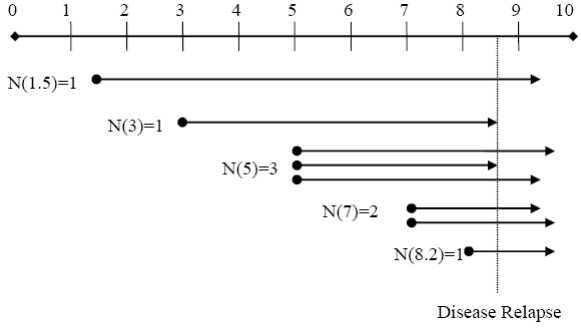

Figure 1. Disease progression diagram

For an individual in the population let $N(t)$ denote the number of carcinogenic cells occurring at time t and C_l, $(l = 1, ..., N^*)$ denote the random time for the l-th carcinogenic cell to produce a detectable cancer mass, C_l are independent and identically distributed with a common distribution function

$$F(y) = 1 - S(y) \tag{2.5}$$

where $N^* = \int_0^y N(t)dt$, represents the total number of active carcinogenic cells that have occurred before relapse at $Y = y$. Note, here if the patient is cured then no carcinogenic cells occurring, that is $N^* = 0$. In the promotion time model, N is assumed to be independent of t and has a Poisson distribution at the beginning of the study, in our model we propose to have $N(t)$ changed over time so that $N(t)$ will follow the non-homogeneous Poisson process with mean $\lambda(t)$.

Theorem 2.1. *If $N(t)$, $t > 0$ is a Poisson process with mean $\lambda(t)$, then $N^* = \int_0^y N(t)dt$ is a Poisson random variable with parameter $\Lambda(y) = \int_0^y \lambda(t)dt$. i.e. $P(N^* = k) = [\Lambda(y)]^k e^{-\Lambda(y)}/k!$.*

Moreover, for $t \in (0,\ y)$, the conditional distribution of the exact time of the occur of an active carcinogenic cells given $N^*(>0)$ are independent and identically with probability density function $g(t) = \lambda(t)/\int_0^y \lambda(t)dt = \frac{\lambda(t)}{\Lambda(y)}\ ,\ \ t \in (0, y)$.

Given N^*, the random variable C_l is assumed to be i.i.d with a common distribution function $F(y) = 1 - S(y)$, which is independent of N^*. The conditional population survival function given $N^*(>0)$ can then be derived as

$$\begin{aligned}\check{S}(y) &= P(Y > y|N^*)\\ &= P(\text{no carcinogenic cells by time } y \text{ given } N^*)\\ &= \prod_{i=1}^{N^*}\left\{\int_0^y \frac{\lambda_i(t)(1-F(y-t))}{\int_0^y \lambda_i(\xi)d\xi}dt\right\}\\ &= \left\{\int_0^y g(t)S(y-t)dt\right\}^{N^*}. \qquad (2.6)\end{aligned}$$

Note, N^* is not observed (latent) variable in the model formulation. Thus, summing (2.6) out N^*, one can obtain the unconditional population survival function as

$$\begin{aligned}S_P(y) &= P(\text{no cancer cells by time } y)\\ &= P(N^* = 0) + P(C_1 > y,\ C_2 > y\ ,...,C_{N^*} > y,\ N^* \geq 1)\\ &= \exp(-\Lambda(y)) + \sum_{k=1}^{\infty}\left\{\left[\int_0^y g(t)S(y-t)dt\right]^k \times \frac{(\Lambda(y))^k \exp(-\Lambda(y))}{k!}\right\}\\ &= \exp\left[\int_0^y -\lambda(t)F(y-t)dt\right]. \qquad (2.7)\end{aligned}$$

We emphasize here that population survival function is the sum of the cured ($N^* = 0$) and non-cured ($N^* > 0$) patients. The cure fraction is thus given by

$$S_P(\infty) = \exp\left\{-\lim_{y\to\infty}\int_0^y \lambda(t)F(y-t)dt\right\}, \qquad (2.8)$$

if the integral in (2.8) is bounded then the survival function has a non-zero cure fraction, otherwise the survival function in (2.7) leads to a proper survival function, that is $S_P(\infty) = 0$.

Using the properties of a distribution function $F(t)$ and the fact that $\lambda(t)$ is non-negative, as $y \to \infty$ the population survival function in (2.8) reduces to

$$S_P(\infty) = \exp\left\{-\lim_{y\to\infty}\Lambda(t)\right\}, \qquad (2.9)$$

that's to say, a cure rate model is characterized by a bounded cumulative mean for the number of carcinogenic cells, while a proper survival model is characterized by an unbounded cumulative risk. And hence, this development of the stochastic disease process allows models for both zero and non-zero cure fractions.

The density function corresponding to (2.7) is given by

$$f_P(y) = \frac{d}{dy}F_P(y)$$

$$= \int_0^y \lambda(t) f(y-t)dt \exp\left[-\int_0^y \lambda(t)F(y-t)dt\right] \tag{2.10}$$

where $f(y) = \frac{d}{dy}F(y)$. The hazard function is then given by

$$h_P(y) = \frac{f_P(y)}{S_P(y)} = \int_0^y \lambda(t) f(y-t)dt. \tag{2.11}$$

Since $S_P(y)$ is not a proper survival function when the integral $\int_0^y \lambda(t)F(y-t)dt$ is unbounded, and hence $f_P(y)$ is not a proper probability density function and $h_P(y)$ is not a hazard function corresponding to a probability distribution. However, $f(y)$ is a proper probability density function and $h_P(y)$ is compound of λ, $F(y)$, and $f(y)$. Thus, it has the proportional hazard structure when the covariates modelled through $\lambda(t)$.

The survival function for the non-cured population is given by

$$\begin{aligned} S_1(y) &= P(Y > y | N^* \geq 1) \\ &= P(N^* \geq 1,\ Y > y)/P(N^* \geq 1) \\ &= \frac{\exp\left[-\int_0^y \lambda(t)F(y-t)dt\right] - \exp\left[-\Lambda(y)\right]}{1-\exp\left[-\Lambda(y)\right]}. \end{aligned} \tag{2.12}$$

Note that $S_1(0) = \lim_{y\to 0} S_1(y) = 1$, and $S_1(\infty) = \lim_{y\to\infty} S_1(y) = 0$, that is, $S_1(y)$ is a proper survival function. The probability density function for the non-cured population is given by

$$\begin{aligned} f_1(y) &= -\frac{d}{dy}S_1(y) \\ &= \frac{\exp\left[-\int_0^y \lambda(t)F(y-t)dt\right]}{1-\exp\left[-\Lambda(y)\right]} \int_0^y \lambda(t) f(y-t)dt, \end{aligned} \tag{2.13}$$

and the hazard function for the non-cured population is then given by

$$\begin{aligned} h_1(y) &= \frac{f_1(y)}{S_1(y)} \\ &= \frac{\exp\left[-\int_0^y \lambda(t)F(y-t)dt\right]}{\exp\left[-\int_0^y \lambda(t)F(y-t)dt\right] - \exp\left[-\Lambda(y)\right]} \int_0^y \lambda(t) f(y-t)dt. \end{aligned} \tag{2.14}$$

The hazard function in (2.14) depends on y, then, we can say that $h_1(y)$ does not have a proportional hazard structure. To write $S_P(y)$ in term of the cure fraction θ, one can use the mathematical relationship between the models in (1.1), (2.7) and (2.9), then the model can be written as

$$\begin{aligned} S_P(y) &= \exp\left[-\int_0^y \lambda(t)F(y-t)dt\right] \\ &= \exp\left[-\Lambda(y)\right] + \left\{1-\exp\left[-\Lambda(y)\right]\right\} S_1(y), \end{aligned} \tag{2.15}$$

thus, $S_P(y)$ is *a standard cure rate model* with cure fraction $\theta = \exp\left[-\Lambda(y)\right]$.

To incorporate information from both the longitudinal trajectories $X^*(t)$ and the other potential covariates (time dependent or time fixed) for survival model. With $N(t)$ being the Poisson count, suppose that we want to let the mean $\lambda(t)$ depend on a vector of explanatory variables (longitudinal trajectories and the other

potential covariates), we take log (canonical link) and assume that the transformed mean follows a linear model. Thus, we consider a generalized linear model with link log as

$$Log(\lambda(t)) = \gamma X^*(t) + \delta Z(t),$$

equivalently, the above model can be written as

$$\lambda(t) = \exp\{\gamma X^*(t) + \delta Z(t)\} \tag{2.16}$$

where γ is a $p \times 1$ vector of regression coefficient represents the effects of the marker on the disease risk, and δ is $q \times 1$ vector of regression coefficient corresponding to the other covariates $Z(t)$. Thus $\lambda(t)$ is the conditional mean of $N(t)$ given $X^*(t)$. Entering the covariates in this fashion corresponds to a canonical link in a Poisson generalized linear model, all covariates are assumed to affect survival through their impact on the mean number of metastasis-competent tumor cells over time. The case $\gamma = 0$ implies that the subject-specific marker response is not associated with the number of carcinogenic cells in the body, i.e. we got separate model.

3. Joint likelihood and priors

In this section, we construct the joint likelihood with a specific choice of the longitudinal trajectory function and distribution assumption of the promotion time. For the longitudinal process, we consider the situation where the only coefficients in U_1 and U_2 in model (2.1), are the intercept t, then, with some change in notation, (2.1) can be written as

$$\begin{cases} X_i(t_{ij}) = X_i^*(t_{ij}) + \epsilon_i(t_{ij}) \\ X_i^*(t_{ij}) = a_i + bt_{ij} + \beta U_i(t_{ij}) + W_i(t_{ij}), \end{cases} \tag{3.1}$$

where $X_i(t_{ij})$, $X_i^*(t_{ij})$ denote the observed and the true values of a continuous time-dependent covariates at time t_{ij} respectively, $a_i \sim N(\mu_a, \sigma_a^2)$ are independent random intercept of subject i, b is the average rate of decline of the marker, $U_i(t_{ij})$ is a $(p \times 1)$ vector of the values of p variables for subject i at time t_{ij}, the $1 \times p$ vector of unknown regression parameter β represents the effect of the p variables on the marker, $W_i(t_{ij}) \sim N(0, \Sigma)$ are independent IOU stochastic process with covariance structure Σ given in (2.2), and $\epsilon_i(t_{ij}) \sim N(0, \sigma_\epsilon^2)$ represents deviations due to measurement error. With the above changes, model in (2.16), then can be written as

$$\lambda_i(t) = \exp\{\gamma\,[a_i + bt + \beta U_i(t) + W_i(t)] + \delta Z_i(t)\}\,. \tag{3.2}$$

We will use Bayesian approach in our modeling, focusing on the estimation of the joint posterior density of all unknown model parameters $\Omega = \{\mu_a, \sigma_a^2, b, \beta, \gamma, \delta, \sigma_e^2, \alpha, \sigma^2, \pi\}$, where π is the promotion time parameter. In the proposed joint model the observed data is given by $D_{obs} = \{n, M, X, Y, \Delta, U, Z\}$ and the complete data is given by $D = D_{obs} \cup \{N^*, a_1, ..., a_n, W_1(t), ..., W_1(t)\}$, where $M = \sum_{i=1}^{n} n_i$.

The joint posterior density of the parameters depends on their prior density and likelihood assumptions, we will specify these assumptions as in model (3.1). We use the notation $[\cdot]$ and $[\cdot \mid \cdot]$ to denote marginal and conditional densities respectively. For the likelihood function, we assume:

(1) The data from different subjects are independent.

(2) For each subject i, given all the unknown parameters in Ω and covariates (U_i, Z_i), his longitudinal data is independent of his survival data.
(3) For each subject i, given $\{X_i^*(t_{ij}),\ j = 1, ..., n_i\}$, $\{X_i(t_{ij})\}_{j=1}^{n_i}$ are independent and $X_i(t_{ij})$ has normal distribution $N(X_i^*(t_{ij}), \sigma_e^2)$.

Thus, the contribution of subject i to the conditional likelihood is

$$
\begin{aligned}
&[X_i(t_{ij}),\ (y_i, \Delta_i) \mid \Omega,\ X_i,\ Z_i] \\
&= [X_i(t_{ij}) \mid \Omega,\ X_i,\ Z_i]\,[y_i, \Delta_i \mid \Omega,\ X_i,\ Z_i] \\
&= \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma_e^2}} \times \exp\left\{ -\frac{(X_i(t_{ij}) - (a_i + bt + \beta U_i(t) + W_i(t)))^2}{2\sigma_e^2} \right\} \qquad (3.3) \\
&\quad \times (S_i(y_i))^{N_i^* \Delta_i} (N_i^* f_i(y_i))^{\Delta_i} \times \frac{\Lambda_i(y_i)^{N_i^*} \exp(-\Lambda_i(y_i))}{N_i^*!}.
\end{aligned}
$$

The likelihood function for the joint model involves two components. The first component involves the longitudinal process denoted by L_1. The second component involves the likelihood function of the time-to-event variable Y, denoted by L_2. Then the likelihood function for the joint model will be the product of L_1 and L_2.

Given the parameters $\Omega_1 = \left\{b, \mu_a, \sigma_a^2, \beta, \alpha, \sigma^2, \sigma_e^2\right\}$. From (3.3) and the longitudinal model assumptions (model (3.1)), then L_1 can be defined as:

$$
\begin{aligned}
L_1(\Omega_1) &= \prod_{i=1}^{n} \{ \prod_{j=1}^{n_i} \left[Y_i(t_{ij}) \mid a_i,\ b,\ W_i(t),\ \beta,\ \sigma_e^2\ X_i \right] \} \\
&\quad \times \left[W_i(t) \mid \alpha,\ \sigma^2 \right] \left[a_i \mid \mu_a,\ \sigma_a^2 \right] \\
&= \prod_{i=1}^{n} \left[\left(\prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left\{ -\frac{(X_i(t_{ij}) - (a_i + bt + \beta U_i(t) + W_i(t)))^2}{2\sigma_e^2} \right\} \right) \right. \qquad (3.4) \\
&\quad \left. \times \frac{\Sigma_i^{-\frac{1}{2}}}{2\pi} \exp\left(-\frac{W_i^T \Sigma_i^{-1} W_i}{2} \right) \times \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left(-\frac{(a_i - \mu_a)^2}{2\sigma_a^2} \right) \right].
\end{aligned}
$$

To complete the second piece of the joint likelihood, we assume that the promotion time of an active metastasis-competent tumor cell independent comes from a common semi-parametric exponential distribution [22]. In this function, we partition the time scale $y_i,\ i = 1, ..., n$ into J intervals, i.e. $0 < s_1 < s_2 < \cdots < s_J,\ s_J > y_i$ for all i. Thus we have J intervals $(0, s_1]$, $(s_1, s_2]$, ..., $(s_{J-1}, s_J]$, we thus assume that the hazard for $F(y)$ is constant and equal to π_j for the jth interval, $j = 1, ..., J$. Then, this function is given by

$$
F(y) = 1 - \exp\left\{ -\pi_j (y - s_{j-1}) - \sum_{q=1}^{j-1} \pi_q (s_q - s_{q-1}) \right\}. \qquad (3.5)
$$

By substituting function (3.5) into model (2.6), the conditional survival function of an active carcinogenic cell to become detectable tumor at time y_i given N^*, can

be derived as

$$
\begin{aligned}
\check{S}(y_i) &= \int_0^{y_i} g_i(t) S(y_i - t) dt \\
&= \frac{1}{\Lambda_i(y_i)} \int_0^{y_i} \lambda_i(t)(1 - F(y_i - t)) dt \\
&= \frac{1}{\Lambda_i(y_i)} \int_0^{y_i} \lambda_i(t) \exp\left\{ -\pi_j \left(y_i - t - s_{j-1}\right) - \sum_{q=1}^{j-1} \pi_q \left(s_q - s_{q-1}\right) \right\} dt \\
&= \frac{1}{\Lambda_i(y_i)} \int_0^{y_i} \exp\left\{\gamma \left[a_i + bt + \beta X_i(t) + W_i(t)\right] + \delta Z_i(t)\right\} \\
&\quad \times \exp\left\{ -\pi_j \left(y_i - t - s_{j-1}\right) - \sum_{q=1}^{j-1} \pi_q \left(s_q - s_{q-1}\right) \right\} dt \\
&= \frac{1}{\Lambda_i(y_i)} \left(\sum_{k=1}^{j} I(y_i > s_{k-1}) \int_{s_{k-1}}^{r} \exp\left\{\gamma \left[a_i + bt + \beta X_i(t) + W_i(t)\right] + \delta Z_i(t)\right\} \right. \\
&\quad \left. \times \exp\left\{ -\pi_k \left(r - t - s_{k-1}\right) - \sum_{q=1}^{k-1} \pi_q \left(s_q - s_{q-1}\right) \right\} dt \right) \\
&= \frac{\exp\left(\gamma a_i\right)}{\Lambda_i(y_i)} \Big(\sum_{k=1}^{j} \exp\left\{ -\pi_k \left(r - s_{k-1}\right) - \sum_{q=1}^{k-1} \pi_q \left(s_q - s_{q-1}\right) \right\} \\
&\quad \times I(y_i > s_{k-1}) \xi(\varsigma) \Big)
\end{aligned}
\tag{3.6}
$$

where j is the interval index such that $y_i \in (s_{j-1},\ s_j]$, $r = \min(y_i,\ s_k)$ and

$$
\xi(\varsigma) = \int_{s_{k-1}}^{r} \exp\left\{\gamma \left[a_i + bt + \beta X_i(t) + W_i(t)\right] + \delta Z_i(t) + \pi_k t\right\} dt.
$$

Information about the continuous stochastic process $W_i(t)$ is needed to calculate $\xi(\varsigma)$. We approximate the continuous function $W_i(t)$ by its value at a finite set of i_w grid points $(t_{i1}^w,\ t_{i2}^w, ..., t_{ii_w}^w)$ in order to facilitate the estimation of all parameters in the joint model. The i_w grid points are chosen to contain all the time points where marker measurements is taken for subject i, since the value of $W_i(t)$ at these points are used in the longitudinal model and needed to be estimated, also we choose the grid points so that the maximum of $\{t_{ij}^w - t_{ij-1}^w,\ \ j = 1, ..., i_w\}$ (assuming $t_{i0}^w = 0$) is very small and $W_i(t)$ can be considered as constant over the interval $\left(t_{ij-1}^w,\ t_{ij}^w\right]$. Further, we assume also that the time dependent covariates (if there any) are constant over the same interval. Since we already partition the scalar time y_i into J intervals then the i_w grid points will be considered only in one interval of J, so that in each grid i_w interval we will assume $t_{i0}^w = s_{k-1},\ k = 1, ..., J$. Thus $\xi(\varsigma)$ can be evaluated as

$$
\xi(\varsigma) = \sum_{l=1}^{l_i} \int_{t'_{l-1}}^{l} \exp\left\{\gamma \left(bt + \beta X_i(t) + W_i(t)\right) + \delta Z_i(t) + \pi_k t\right\} dt
$$

$$
\begin{aligned}
&- \int_{t'_{l_i}}^{r} \exp\left\{\gamma\left(bt+\beta X_i(t)+W_i(t)\right)+\delta Z_i(t)+\pi_k t\right\} dt \\
&= \sum_{l=1}^{l_i}\sum_{m=1}^{M_i(l)} \int_{t''_{l(m-1)}}^{t''_{lm}} \exp\left\{\gamma\left(bt+\beta X_i(t)+W_i(t)\right)+\delta Z_i(t)+\pi_k t\right\} dt \\
&\quad - \sum_{m=1}^{M_i(r)} \int_{t''_{r(m-1)}}^{t''_{rm}} \exp\left\{\gamma\left(bt+\beta X_i(t)+W_i(t)\right)+\delta Z_i(t)+\pi_k t\right\} dt \\
&= \sum_{l=1}^{l_i}\sum_{m=1}^{M_i(l)} \exp\left\{\gamma\left(\beta X_i(t''_{lm})+W_i(t''_{lm})\right)+\delta Z_i(t''_{lm})\right\} \\
&\quad \times \frac{\exp\left(\gamma b+\pi_k\right) t''_{lm} - \exp\left(\gamma b+\pi_k\right) t''_{l(m-1)}}{\left(\gamma b+\pi_k\right)} \\
&\quad - \sum_{m=1}^{M_i(r)} \exp\left\{\gamma\left(\beta X_i(t''_{rm})+W_i(t''_{rm})\right)+\delta Z_i(t''_{rm})\right\} \\
&\quad \times \frac{\exp\left(\gamma b+\pi_k\right) t''_{rm} - \exp\left(\gamma b+\pi_k\right) t''_{r(m-1)}}{\left(\gamma b+\pi_k\right)},
\end{aligned}
$$

where $r=\min(y_i,s_k)$, $l_i=\max\{l:\ t'_l\le r\}$, for $l=1,...,l_i$, $t''_{l0}=t'_{l-1}$, $t''_{lM_i(l)}=t'_l$ and $\left(t''_{l1},\ t''_{l2},...,\ t''_{l(M(l)-1)}\right)$ all are grid points ordered in interval $\left(t'_{l-1},\ t'_l\right)$, for subject i ; $\left(\ t''_{r1},\ t''_{r2},\ \ldots\ ,t''_{r(M_i(r)-1)}\right)$ all are grid points ordered in interval $(t'_{li},\ r)$ for subject i, $t''_{r0}=t'_{li}$, $t''_{l_iM_i(r)}=r$, and $t'_0=s_{k-1}$.

Given N_i^* the conditional distribution function $\tilde{F}_i(y_i)$ for an active carcinogenic cells to become a detectable tumor cells at time y_i is given by

$$\tilde{F}(y_i)=1-\check{S}(y_i) \tag{3.7}$$

also the conditional density function is given by

$$\tilde{f}(y_i)=\frac{d}{dy_i}\tilde{F}(y_i)=\frac{d}{dy_i}(1-\check{S}_i(y_i))=\pi_j\check{S}_i(y_i). \tag{3.8}$$

In the same manner, the cumulative rate $\Lambda_i(y_i)$ is given by

$$
\begin{aligned}
\Lambda_i(y_i) &= \int_0^{y_i} \lambda_i(t)dt \\
&= \sum_{k=1}^{k_i}\sum_{j=1}^{J_i(k)} \exp\left\{\gamma\left(a_i+bt^i_{kj}+\beta X_i(t^i_{kj})+W_i(t^i_{kj})\right)+\delta Z_i(t^i_{kj})\right\} \\
&\quad \times \left\{\exp\left(\gamma b t^i_{kj}\right)-\exp\left(\gamma b t^i_{k(j-1)}\right)\right\}/\gamma b \\
&\quad - \sum_{j=1}^{J_i(y_i)} \exp\left\{\gamma\left(a_i+bt^i_{y_{ii}j}+\beta X_i(t^i_{y_{ii}j})+W_i(t^i_{y_{ii}j})\right)+\delta Z_i(t^i_{y_{ii}j})\right\}
\end{aligned} \tag{3.9}
$$

$$\times \left\{ \frac{\exp\left(\gamma b t^i_{y_i j}\right) - \exp\left(\gamma b t^i_{y_i (j-1)}\right)}{\gamma b} \right\}.$$

Given the parameters $\Omega_2 = \{\gamma, \delta, \pi_1, ..., \pi_J\}$. From (3.3) and the survival model assumptions, then the second component of the likelihood function L_2 can be derived as

$$\begin{aligned} L_2(\Omega_2) &= P\left(y_i;\ \Delta_i \mid a_i,\ b,\ W_i(t),\ \beta,\ X_i,\ Z_i,\ N_i^*\right) P\left(N_i^* \mid \gamma,\ \delta\right) \\ &= \prod_{i=1}^{n} \prod_{j=1}^{J} \left(\check{S}_i(y_i)\right)^{(N_i^* - \Delta_i)\Delta_{ij}} \left(N_i^* \tilde{f}(y_i)\right)^{\Delta_i \Delta_{ij}} \\ &\quad \times \exp\left\{ \sum_{i=1}^{n} N_i^* \log\left(\Lambda_i(y_i)\right) - \log\left(N_i^*!\right) - \Lambda_i(y_i) \right\}, \end{aligned} \tag{3.10}$$

where $\check{S}_i(y_i)$, $\tilde{f}(y_i)$ and $\Lambda_i(y_i)$ are given in (3.6), (3.8) and (**??**) respectively, Δ_{ij} censored indicator equal one if the ith subject fails in the jth interval and zero otherwise.

The prior specification for $\Omega = \Omega_1 \cup \Omega_2$ are given jointly as

$$[\Omega] = [b]\,[\mu_a]\left[\sigma_a^2\right][\beta]\,[\alpha]\left[\sigma^2\right]\left[\sigma_e^2\right][\gamma]\,[\delta]\,[\pi_j], \tag{3.11}$$

and hence, the joint likelihood of the complete data is given by

$$L(\Omega) = L_1(\Omega_1) L_2(\Omega_2)\,[\Omega] \tag{3.12}$$

We take $b, \mu_a, \beta, \gamma, \delta$ to have normal priors. For $\sigma_a^2, \sigma^2, \sigma_e^2$ we take inverse gamma priors. The corresponding prior for α has a scaled $\mathbf{F}$ distribution $\mathbf{F}(r, s)$ if $r \neq s$; otherwise is a $\mathbf{F}$ distribution $\mathbf{F}(r, r)$. Finally, we take independent gamma prior for π as follows:

$$[\pi] \propto \prod_{j=1}^{J} \pi_j^{\zeta_0 - 1} \exp(\tau_0 \pi_j),$$

where ζ_0 and τ_0 are pre-specified hyperparameters. The choices of these priors are based on the joint posterior distributions. (See the Appendix)

4. Bayesian model assessment

To assess the model fit and compare different models, we calculate the Conditional Predictive Ordinate (CPO), Gelfand *et al.*[16], and the Deviance Information Criterion (DIC) recently proposed by Spiegelhalter *et al.* [33] where the formulas given by

$$CPO_i = \left(\int \frac{1}{f(X_i^*, Y_i, \Delta_i \mid \phi, U_i, Z_i)} [\phi \mid D] d\phi \right)^{-1} \tag{4.1}$$

where $[\phi \mid D]$ is the posterior density of ϕ based on the data including all subjects.

Using (4.1) a Monte Carlo method presented in Chen *et al.* [7] is readily used for computing CPO_i if $f(X_i^*, Y_i, \Delta_i \mid \phi, U_i, Z_i)$ can be evaluated for each ϕ. However, due to the complexity of the joint model, an analytical evaluation of $f(X_i^*, Y_i, \Delta_i \mid$

ϕ,U_i,Z_i) does not appear possible. Therefore, an alternative Monte Carlo approximate of CPO_i will be used, which is given by

$$\widehat{CPO_i} = \left(\frac{1}{M} \sum_{m=1}^{M} \frac{1}{L_i(\phi_{[m]})} \right)^{-1} \tag{4.2}$$

models with greater $\sum_{i=1}^{n} \log(CPO_i)$ indicate a better fit, and

$$DIC = -\frac{4}{M} \sum_{m=1}^{M} \log L(\phi_{[m]}) + 2\log(\overline{\phi_{[m]}}), \tag{4.3}$$

the smallest the DIC, the better the fit of the model.

5. Sampling methods and simulation study

In this section, we will evaluate the performance of the proposed joint model by conducting a simulation study. We investigate how will the population parameters can be estimated in terms of bias and converge rate, and compare these results to that of the separate modeling approach by applying methods of MCMC sampler. Also we study how the following factor affect the performance of the joint model: Censoring rate and prior information for the parameters.

5.1. Simulation design

To illustrate our joint semiparametric model, we setup our simulation study represent a randomize clinical trial, in which $n = 100$ subjects are randomized. Each longitudinal marker in model (3.1), $X_i(t_{ij})$, $i = 1, ..., n;\ j = 1, ..., n_i,$ was simulated as the sum of the trajectory function $X_i^*(t_{ij})$ and the error terms $\epsilon_i(t_{ij})$, each subject has its observed longitudinal measured $n_i = 10$ at time points $t_1 = 0.1, ..., t_{10} = 1$, until the relapse or the end of the study. For the survival data, we consider a model in the presence of cure; that is we took the mean of the Poisson process at time t as in (3.2) to be for $i = 1, ..., 100,$ where Z_i is a binary baseline covariates with half of the subjects having one and the other half having zero, and the promotion time was considered as in (3.5) with $J = 1$. This setup leads to a cure rate structure for the survival time in model (3.6). We will modeled the longitudinal data and survival data separately, i.e. for longitudinal data, we will use model (3.1) and for survival data we will use model (3.2) with $\gamma = 0$ along with model (3.6), then the maximum likelihood in (3.4) and (3.10) were estimated to get an initial estimate of the population parameters $\Omega = \{\mu_a, \sigma_a^2, b, \beta, \gamma, \delta, \sigma_e^2, \alpha, \sigma^2, \pi_1\}$ say $\Omega^{(0)} = \{\mu_a^{(0)}, \left(\sigma_a^2\right)^{(0)}, b^{(0)}, \beta^{(0)}, \gamma^{(0)}, \delta^{(0)}, \left(\sigma_e^2\right)^{(0)}, \alpha^{(0)}, \left(\sigma^2\right)^{(0)}, (\pi_1)^{(0)}\}$ and use them as initial values in MCMC sampler.

With the Bayesian approach, all estimates and inferences are made on the posterior distribution of the parameters of interest. Combining with the likelihood based on the available data, prior distributions are used to derive the posterior density of the parameters. To implementing the MCMC sampler algorithms for our joint modeling approach, based on the joint posterior distribution of the parameters in

section 4, the full conditional distribution of the parameters are derived (see Appendix). We note that the the full conditional distributions of the parameters σ_e^2, μ_a , σ_a^2 and σ^2 that appearing only in the longitudinal model are a product of its prior density and some standard distribution which are conjugate priors for these parameters. While the conditional distribution of the parameters b, $(\beta_1, ..., \beta_p)$, if the contributions from the survival data are ignored, then the normal distribution are conjugate priors, if the contributions from the survival data are not ignored, then we will use it as a proposed density in ARMS sampler. The main difficulty which we will meet in the prior distributions is that when no standard form appears in the posterior distribution. In general, we do not have performance in choosing priors for the parameters α, γ, δ since their full conditional densities have no conjugate priors. One may use normal priors for γ, δ since they take values belong to the real line $\mathbb{R}$. For the IOU stochastic process parameter α, gamma and inverse gamma distributions are potential choices as priors since it takes only positive values.

Throughout this section, the true values of the population parameters, which are used to generate the 100 data sets are $b = -3.5$, $\beta = 1$, $\mu_a = 4.0$, $\sigma_a^2 = 0.02$, $\alpha = 0.138$, $\sigma^2 = 0.12$, $\sigma_e^2 = 0.05$, $\gamma = -1$, $\delta = 2.4$, and $\pi_1 = 0.05$. All the parameters were assumed independent a proiori and assigned non-informative priors, so we choose $b \sim N(-4.00, 1.00)$; $\beta \sim N(1.5, 0.50)$; $\mu_a \sim N(4.00, 1.00)$; $\sigma_a^2 \sim IG(2.00, 0.01)$; $\alpha \sim F(1.5, 1.5)$; $\sigma^2 \sim IG(1.00, 0.02)$; $\sigma_e^2 \sim IG(2.00, 0.01)$; $\gamma \sim N(-1.5, 1.0)$; $\delta \sim N(3.0, 1.0)$ and $\pi_k \sim G(0.02, 1.0)$.

For the parameters $\left\{\mu_a, \sigma_a^2, \sigma^2, \sigma_e^2, \pi_1\right\}$ drawing random variates from their full conditional distribution is straight forward, therefore, we will use the full conditional density as a proposal density in Gibbs sampler algorithm, and in sampling process each updating step for these parameters, a new draw from the full conditional density is always accepted. We perform this algorithm for each parameter 2,000 Gibbs samples after 1,000 burn-in. The histogram, the time series plots of one sequence of Gibbs samples for different number of iterations and the average number of these iterations for the parameter μ_a are presented in Figure 2.

For the parameters $\{b, \beta\}$ one can not draw a random variate from these densities directly due to the terms from the time-to-event data. For each one of these parameters, we use the Metropolis-Hastings (M-H) algorithm to obtain the update in the Gibbs sampling sequence. With the Gibbs algorithm, a proposal density is required to draw a random variates and to be compared with the full conditional density at this random variate and at the current value of the parameter. So we will use the standard density, which we got from the contribution of the longitudinal data and priors as a proposal density. The histogram, the time series plots of one sequence of Gibbs samples for different number of iterations and the average number of these iterations for the parameter β are presented in Figure 3.

For the parameters $\{\alpha, \gamma, \delta_1 ..., \delta_p\}$ is not follow any standard distribution, it is just an algebraic expression which come from the contribution of the longitudinal and time-to-event data, so that, for such parameters, one can not draw random variates from their full conditional densities. For each of these parameters we propose using a normal density as a proposal density, and then the Adaptive Rejection Metropolis Sampling (ARMS) [17] within Gibbs sampling will be used by considering $f(x) = q(\theta|D)$, and then constructing a sampling distribution function $g(x)$ for which

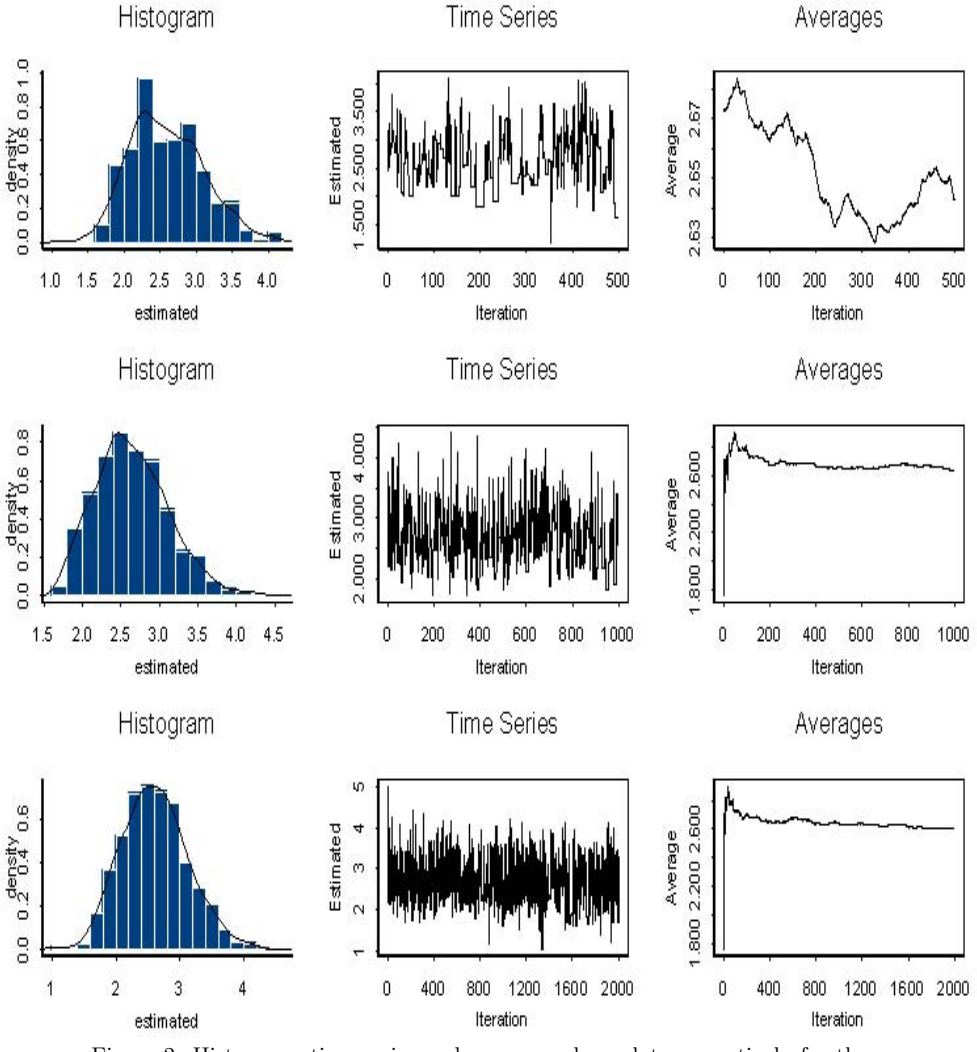

Figure 2. Histogram, time series and average values plots respectively for the parameter values μ_a at 500, 1000, and 2000 iterations respectively, using Gibbs sampler.

samples can be readily drawn. The histogram, the time series plots of one sequence of Gibbs samples for different number of iterations and the average number of these iterations for the parameter δ are presented in Figure 4. Figures 2, 3 and 4 show that these sequences are mix well and converge within 2,000 iterations after 1,000 iterations are burn-in.

Given a set of values of parameters in Ω, the data for subject i can be generated as follows:

(1) Simulate the discrete IOU process $W_i = (W_{ij} = W_i(t_j))$ by drawing a multivariate random variates from the normal distribution $N_{10}(0, \Sigma)$, where Σ

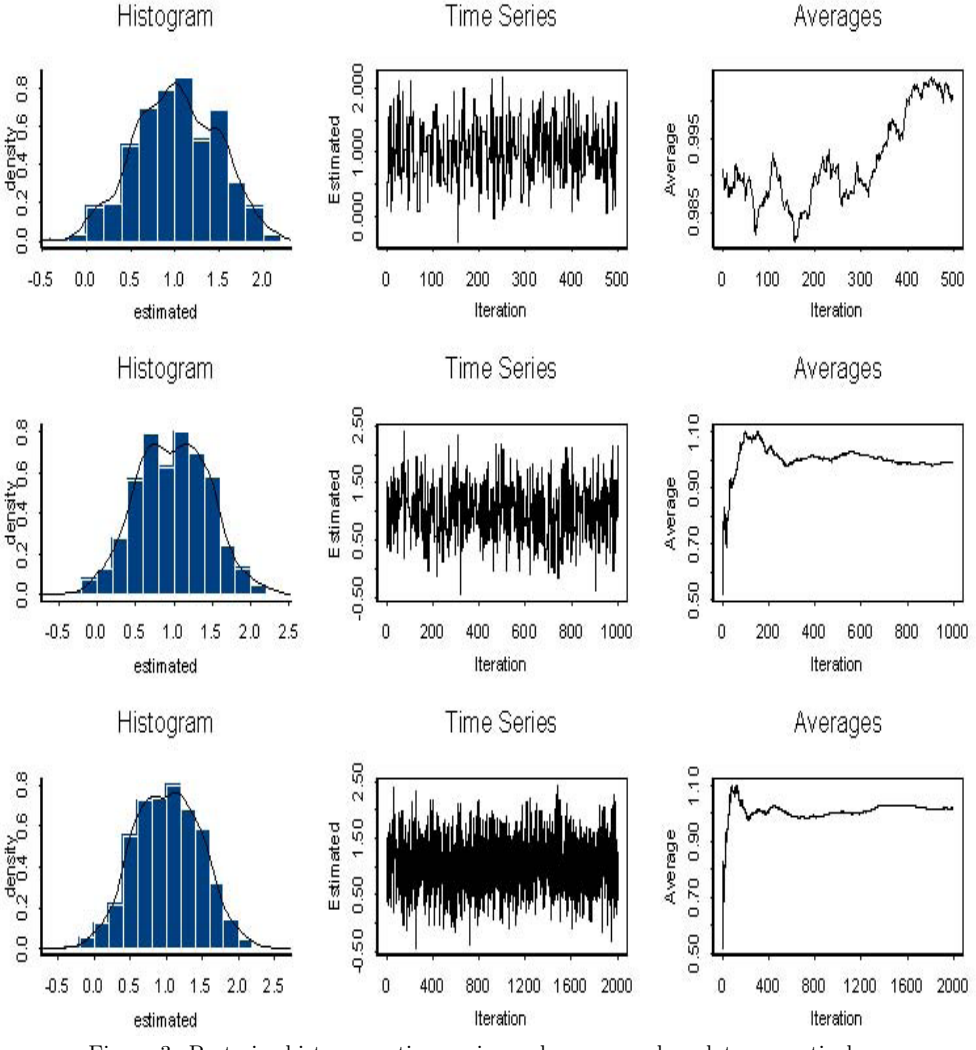

Figure 3. Posterior histogram, time series and average value plots respectively for the parameter values β at 500, 1000, and 2000 iterations respectively, using MH sampler.

is the variance covariance matrix defined by (2.4). Simulate random intercept a_i by drawing a random variate from the univariate normal distribution $N(\mu_a, \sigma_a^2)$. Simulate the true longitudinal measurements X_i^*, by the lower part of model (3.1). Simulate the measurements error ϵ_i from the univariate normal distribution $N(0, \sigma_e^2)$. Simulate the observed longitudinal measurements X_i by adding a measurement error to the true longitudinal measurements as in upper part of model (3.1).

(2) Simulate the failure time Y_i under model (3.2), and suppose that a subject has not contracted an active carcinogenic cells up to time t_j, then the

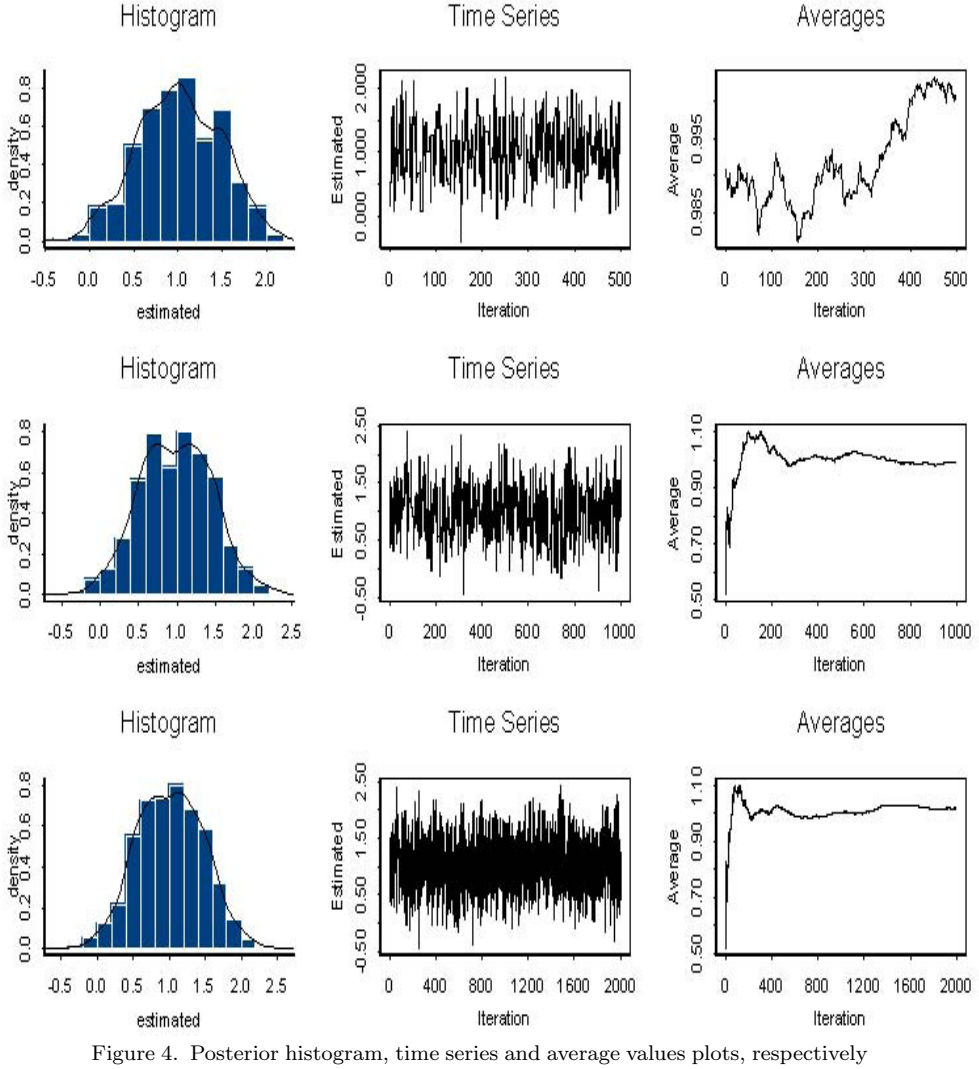

Figure 4. Posterior histogram, time series and average values plots, respectively for the parameter values δ at 500, 1000, and 2000 iterations respectively, using ARMS sampler.

probability that he will develop carcinogenic tumor cells during the time period $(t_j, t_{j+1}]$ is approximately $p_i = \lambda_i(t \mid X_i^*(s), Z_i, s \leq t) \times (t_{j+1} - t_j)$. We draw a random variable $U_1 \sim U(0,1)$. If $p_i < U_1$, we say that he develops carcinogenic cells over time interval $(t_j, t_{j+1}]$, and draw a random variate $U_2 \sim U(0,1)$ and define his survival time as $Y_i = t_j + U_2 \times (t_{j+1} - t_j)$, otherwise he is still carcinogenic cells free at time t_{j+1}, in this case set $\Delta_i = 0$. We continue this process until either he develop carcinogenic cells or the time reached the maximum follow-up period. Following this process, generating of censoring indicator Δ_i. Also by the above techniques one can

controls the cure and censoring rates. In this simulation study, we choose moderate cure-rate $(15-30)\%$, and moderate censoring $(30-50)\%$. Simulate the Poisson process $N_i(t)$ by drawing a random variates from $\text{Pois}(\Lambda_i(y_i)\ \check{S}(y_i)) + \Delta_i$.

(3) Fit the generated data with proposed joint modelling approach model (3.6). For the purpose of comparison, also model the longitudinal measurements and the survival data separately to obtain CPO_i, DIC, and corresponding parameter estimates.

We repeat the above process 100 times and pool the results to evaluate the overall performance of the estimates of the parameters by evaluating the summary statistics.

5.2. Numerical results

With the initial values of the parameters for which the data are generated considered as the truth values of the parameters, estimate Monte Carlo Summary statistics, Monte Carlo Standard Deviation (MCSD), Mean Squared Error (MSE), 95% Confidence Converge Rate (CCR), and Bias in Percentage Terms (BPT) are presented in Table 1, where, MCE stand for Monte Carlo Error and it can be evaluated as follows: In our simulate study we used 100 data replications, thus the resulting estimates are subject to sampling variation (Monte Carlo Error), this variation for the point estimate can be calculated as $\hat{p} = MCSD/\sqrt{100}$, the MCE then can be found by $MCE = \sqrt{\frac{\hat{p}(1-\hat{p})}{100}}$.

The results in Table 1 assert the convergence of the Markov Chain and the samplers reached the convergence after 2,000 iterations after 1,000 iterations are burn-in. Posterior means, posterior standard deviations, Bias as percent of true parameter and 95% highest posterior density intervals for each parameter in the joint and separate models, are represented in Table 2. These summarize the results for the parameters in the longitudinal model and in the survival model. The estimates of all parameters from the joint modelling analysis are quite accurate and efficient. The estimates are close to the true values of the parameters and have good coverage rates. The small biases of the estimates are due to Monte Carlo simulation error. Compared to the separate model, the joint model results in improved estimates almost for all parameters in both longitudinal and survival model.

By using $M = 1000$ for model assessment for measuring the $LPML$ statistic, also to assert our assessment, DIC were calculated for different models, the results are described in Table 4. A1–A3 models in this table referred to model (3.6) along with some changes in model (3.2), that is, model (A1) referred to model (3.1) include the IOU term (α is finite), model (A2) referred to mixed effects model i.e. model (3.1) excluded the IOU term, in model (A3) the Brownian motion term replaced by the IOU term in model (3.1) (α is infinitely large). Model (B1) referred to model (3.1), and model (B2) referred to model (3.6) with ($\gamma = 0$) in model (3.2). We observe that the joint model that include the IOU stochastic process and the Joint Model corresponds to Brownian Motion give a better fit to the data than the one excluding the IOU term and separately. In other words, the longitudinal model with the IOU stochastic process or Brownian motion (α is finite or infinitely large) yields a superior fit than the model with the random effects, also comparing the values

Table 1. Monte Carlo Summary statistics of the parameter estimates.

Parameter	True Value	Estimated Value	MCSD	MSE	95% CCR	BP	MCE
b	-3.500	-3.498	0.031	8.236×10^{-4}	95%	-0.057%	6×10^{-3}
μ_a	4.000	4.001	0.019	3.450×10^{-4}	98%	0.025%	4×10^{-3}
σ_a^2	0.020	0.020	0.005	2.471×10^{-5}	99%	-0.001%	2×10^{-3}
α	0.138	1.400	0.967	0.928	93%	1.450%	3×10^{-2}
σ^2	0.120	0.119	0.0823	6.622×10^{-3}	96%	-0.833%	9×10^{-3}
β	1.000	0.997	0.041	1.688×10^{-3}	95%	$-.274\%$	6×10^{-3}
σ_e^2	0.050	0.050	0.008	6.855×10^{-5}	94%	0.140%	3×10^{-3}
γ	-1.000	-1.003	0.117	1.469×10^{-4}	97%	0.291%	1×10^{-2}
δ	2.400	2.401	0.259	0.0701	94%	0.042%	2×10^{-2}

Table 2. Posterior estimates from joint and separate models for $J = 1$.

	Joint Models				Separate Models			
Parameter	mean	SD	Bias%	95% C.I.	Mean	SD	Bias%	95% C.I.
b	-3.490	.102	-1.0%	(-3.69, -3.290)	-3.489	.110	-1.1%	(-3.71, -3.273)
β	0.997	.028	0.3%	(0.942, 1.052)	0.991	.090	0.9%	(0.815, 1.167)
μ_a	4.001	.013	-0.1%	(3.975, 4.026)	4.001	.013	-0.1%	(3.975, 4.026)
σ_a^2	0.019	.011	0.1%	(0.000, 0.041)	0.019	.011	0.1%	(0.000, 0.041)
α	1.530	.283	-15.%	(0.850, 1.960)	1.570	.352	-19.%	(0.724, 2.104)
σ^2	0.119	.016	0.1%	(0.088, 0.150)	0.119	.016	0.1%	(0.088, 0.150)
σ_e^2	0.051	.011	-0.1%	(0.294, 0.073)	0.051	.011	-0.1%	(0.294, 0.073)
γ	-1.008	.080	0.8%	(-1.17, -0.851)	-1.013	.127	1.3%	(-1.26, -0.76)
δ	2.405	.051	-0.5%	(2.305, 2.505)	2.391	.088	0.9%	(2.219, 2.563)
π_1	0.052	.045	-0.2%	(0.000, 0.140)	0.094	.067	-4.4%	(0.000, 0.225)

of $LPML$ and DIC statistics for joint model and the separate models, the results indicate that the joint cure rate model appear to provide a more adequate fit to the simulated data than the separate models.

Table 3. The $LPML$ and DIC statistics for different models.

Model		$LPML$	DIC
(A1) Joint Model IOU included		-728.23	1379.04
(A2) Joint Model IOU excluded		-767.39	1409.20
(A3) Joint Model with Brownian Motion		-729.17	1390.75
Separate Models :	(B1) Survival	-286.91	633.670
	(B2) Longitudinal	-550.72	778.430
	Total	-837.63	1412.10

6. Summary, conclusion and future study

We have proposed a new model for jointly modeling longitudinal and survival data in presence of cure fraction. For the longitudinal process, our model consisting of a combination of mixed effects and an IOU stochastic process. This model is more flexible and plausible than a random effects model since it allows the marker to vary a round a straight line and allows the data to determine the degree of this variation.

Motivated by the promotion time model, a cure rate survival model is proposed by specifying an alternative mechanism for the characteristics of tumor growth. Instead of assuming the carcinogenic cells turn active only at the beginning of the study, we allow the possibility that active carcinogenic cells may occur at any time. Thus, the risk of developing a cancer relapse becomes dynamic over time. Moreover, to control the parametricity and the nonparametricity of the survival function, a common semi-parametric exponential function was assumed for the promotion time of an active metastasis-competent tumor cell. To incorporate information from both the longitudinal trajectories and other covariates, we let all the covariates depend on the rate $\lambda(t)$ through a relation corresponds to a canonical link in a Poisson generalized linear model.

The proposed joint model has several advantages. First, each longitudinal component has its own treatment dependent process $Y_i(t)$. Second, the joint model does not require that all longitudinal response be observed at a given time point for the ith patient. This is an important feature of the model since in cancer trials, it is very difficult to take some measurements at time t. Finally, the proposed joint model is suitable for any type of survival data as long as the data can be thought of as being generated by process of latent risks. Thus the model can be useful for analyzing various types of survival data, also can accommodate a proper and improper survival function.

A Bayesian approach was taken to fit the proposed joint model through a simulation study. The numerical results in the simulation study demonstrate that the joint modelling method results in efficient estimates and good coverage for the population parameters. Also it indicates that when ignoring the association between the longitudinal and the survival data would lead to biased estimates for the most important parameters.

For future works, the proposed joint model will be extended to include multivariate longitudinal and multivariate time-to-event data. However, to induce correlation between failure times, shared frailty will be introduced.

Appendix.

For each of the parameters $\{\sigma_e^2,\ \mu_a\ ,\ \sigma_a^2,\ \sigma^2,\ \pi_j\}$, the full conditional posterior distribution is the product of its prior density and some standard distribution. As an example, we derive the conditional posterior of σ_e^2 as follows:

$$
\begin{aligned}
[\sigma_e^2 \mid \cdot] &\propto \prod_{i=1}^{n}\prod_{j=1}^{n_i}[X_i(t_{ij} \mid a_i,\ b,\ W_i(t),\ U_i(t_{ij})][\sigma_e^2] \\
&\propto \prod_{i=1}^{n}\prod_{j=1}^{n_i}\frac{1}{\sqrt{\sigma_e^2}}\exp\left(-\frac{(X_i(t_{ij})-(a_i+bt_{ij}+\beta U_i(t_{ij})+W_i(t_{ij})))^2}{2\sigma_e^2}\right)[\sigma_e^2]
\end{aligned}
$$

$$\propto IG\left(\alpha_{\sigma_e^2},\ \beta_{\sigma_e^2}\right)[\sigma_e^2],$$

where

$$\alpha_{\sigma_e^2} = \frac{\sum_{i=1}^{n} n_i}{2},$$

and

$$\beta_{\sigma_e^2} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left(X_i(t_{ij}) - (a_i + bt_{ij} + \beta U_i(t_{ij}) + W_i(t_{ij}))\right)^2}{2}.$$

In the same manner we found that the full conditional posterior distributions of $\{\mu_a\ ,\ \sigma_a^2,\ \sigma^2,\ \pi_j\}$ are

$$[\mu_a \mid \cdot] \propto N\left(\frac{\sum_{i=1}^{n} a_i}{n},\ \frac{\sigma_a^2}{n}\right)[\ \mu_a],$$

$$\left[\sigma_a^2 \mid \cdot\right] \propto IG\left(\left(\frac{n}{2} - 1\right),\ \frac{\sum_{i=1}^{n}\left(a_i - \mu_a\right)^2}{2}\right)\left[\ \sigma_a^2\right],$$

$$\left[\sigma^2 \mid \cdot\right] \propto IG\left(\frac{\sum_{i=1}^{n} i_w}{2} - 1,\ \frac{\sum_{i=1}^{n} W_i^T \frac{\Sigma_i^{-1}}{\sigma^2} W_i}{2}\right)\left[\ \sigma^2\right],$$

and

$$[\pi_j \mid \cdot] \propto G\left(\alpha_\pi, \beta_\pi\right),$$

where

$$\alpha_\pi = \sum_{i=1}^{n} \Delta_i I\left(s_{k-1} < y_i \leq s_k\right)$$

and

$$\beta_\pi = \sum_{i=1}^{n} I\left(y_i > s_{k-1}\right)\int_{s_{k-1}}^{r} \exp\left\{\gamma\left[a_i + bt + \beta X_i(t) + W_i(t)\right] + \delta Z_i(t)\right\}.$$

For each of the parameters $\{b, \beta,\ a_i,\ N_i^*\}$, the full conditional posterior distribution is the product of its prior density, some standard distribution and a thired term. As an example, for the average rate of decline of the longitudinal measurements b, the full conditional distribution can be derived as follows:

$$\begin{aligned}
[b \mid \cdot] &\propto \prod_{i=1}^{n}\prod_{j=1}^{n_i}[X_i(t_{ij}) \mid a_i,\ b,\ \beta,\ W_i(t),\ \sigma_e^2][b] \\
&\quad \times \prod_{i=1}^{n}(\Lambda_i(y_i)\tilde{f}_i(y_i))^{\Delta_i}\exp(-\Lambda_i(y_i)(1 - \check{S}_i(y_i))) \\
&\propto \exp\left\{-\frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left(X_i(t_{ij}) - (a_i + bt_{ij} + \beta U_i(t_{ij}) + W_i(t_{ij}))\right)^2}{2\sigma_e^2}\right\}[b] \\
&\quad \times \exp\sum_{i=1}^{n}\{\Delta_i \log(\tilde{f}_i(y_i)) + \Delta_i \log\left(\Lambda_i(y_i)\right) - \Lambda_i(y_i)(1 - \check{S}_i(y_i))\}
\end{aligned}$$

$$\propto \exp\left\{-\frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i} b^2t_{ij}^2 - bt_{ij}\left[X_i(t_{ij}) - a_i - \beta U_i(t_{ij}) - W_i(t_{ij})\right]}{2\sigma_e^2}\right.$$
$$\left. + \frac{\left[X_i(t_{ij}) - a_i - \beta U_i(t_{ij}) - W_i(t_{ij})\right]^2}{2\sigma_e^2}\right\}$$
$$\times \exp\sum_{i=1}^{n}\{\Delta_i \log(\tilde{f}_i(y_i)) + \Delta_i \log\left(\Lambda_i(y_i)\right) - \Lambda_i(y_i)(1-\check{S}_i(y_i))\}[b]$$
$$\propto \exp\left\{-\frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i} t_{ij}^2\left(b - t_{ij}X_i(t_{ij}) - a_i - \beta U_i(t_{ij}) - W_i(t_{ij})\right)^2}{2\sigma_e^2}\right\}[b]$$
$$\times \exp\sum_{i=1}^{n}\{\Delta_i \log(\tilde{f}_i(y_i)) + \Delta_i \log\left(\Lambda_i(y_i)\right) - \Lambda_i(y_i)(1-\check{S}_i(y_i))\}$$
$$\propto N(\mu_b, \sigma_b^2)\exp\sum_{i=1}^{n}\{\Delta_i \log(\tilde{f}_i(y_i)) + \Delta_i \log\left(\Lambda_i(y_i)\right) - \Lambda_i(y_i)(1-\check{S}_i(y_i))\}[b],$$

where,

$$\mu_b = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i} t_{ij}\left(X_i(t_{ij}) - a_i - \beta U_i(t_{ij}) - W_i(t_{ij})\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n_i} t_{ij}^2}$$

and

$$\sigma_b^2 = \frac{\sigma_e^2}{\sum_{i=1}^{n}\sum_{j=1}^{n_i} t_{ij}^2}.$$

In the same way, we found that the full conditional posterior distributions of $\{\beta,$ $a_i,$ $N_i^*\}$ are

$$[\beta_l \mid \cdot] \propto N(m_{\beta_l}, v_{\beta_l})[\ \beta_l]$$
$$\times \exp\sum_{i=1}^{n}\left\{\Delta_i \log\left(\tilde{f}_i(y_i)\right) + \Delta_i \log\left(\Lambda_i(y_i)\right) - \Lambda_i(y_i)\left(1-\check{S}_i(y_i)\right)\right\},$$

where

$$m_{\beta_l} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i} X_{il}(t_{ij})\left(Y_i(t_{ij}) - (a_i + bt_{ij} + \beta(-l)X_i(-l)(t_{ij}) + W_i(t_{ij}))\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n_i} X_{il}^2(t_{ij})},$$

and

$$v_{\beta_l} = \frac{\sigma_e^2}{\sum_{i=1}^{n}\sum_{j=1}^{n_i} X_{il}^2(t_{ij})},$$

$$[a_i \mid \cdot] \propto N(m_a, v_a)\exp\left\{\Delta_i \log\left(\tilde{f}_i(y_i)\right) + \Delta_i \log\left(\Lambda_i(y_i)\right) - \Lambda_i(y_i)\left(1-\check{S}_i(y_i)\right)\right\},$$

where

$$m_a = \frac{\frac{\mu_a}{\sigma_a^2} + \sum_{j=1}^{n_i}\left(Y_i(t_{ij}) - bt_{ij} - \beta X_i(t_{ij}) - W_i(t_{ij})\right)}{\frac{n_i}{\sigma_e^2} + \frac{1}{\sigma_a^2}},$$

and

$$v_a = \frac{1}{\frac{n_i}{\sigma_e^2} + \frac{1}{\sigma_a^2}},$$
$$[N_i^* \mid \cdot] \propto \text{Pois}(\check{S}_i(y_i)\Lambda_i(y_i)) + \Delta_i.$$

For the IOU parameter α, the full conditional density can be derived as:

$$\begin{aligned}
[\alpha \mid \cdot] &\propto \prod_{i=1}^{n} \left[W_i = \left(W_i(t_{i1}^w),\ W_i(t_{i2}^w), ...,\ W_i(t_{ii_w}^w)\right) \mid \alpha,\ \sigma^2\right] [\ \alpha] \\
&\propto \prod_{i=1}^{n} \frac{1}{\Sigma_i^{\frac{1}{2}}} \exp\left(-\frac{W_i^T \Sigma_i^{-1} W_i}{2}\right) [\ \alpha] \\
&\propto \frac{\exp\left(-0.5\sigma^2 \sum_{i=1}^{n} W_i^T \frac{\Sigma_i^{-1}}{\sigma^2} W_i\right)}{\prod_{i=1}^{n} \left|\frac{\Sigma_i}{\sigma^2}\right|^{\frac{1}{2}}} [\ \alpha].
\end{aligned}$$

Finally, for the regression coefficients parameters γ and δ in the joint model, their full conditional densities are given by

$$\begin{aligned}
[\gamma \mid \cdot] &\propto \prod_{i=1}^{n} (\Lambda_i(y_i)\tilde{f}_i(y_i))^{\Delta_i} \exp(-\Lambda_i(y_i)(1 - \check{S}_i(y_i)))\ [\gamma] \\
&\propto \exp \sum_{i=1}^{n} (\Delta_i \log(\Lambda_i(y_i)\tilde{f}_i(y_i)) - \Lambda_i(y_i)(1 - \check{S}_i(y_i)))[\gamma].
\end{aligned}$$

and

$$\begin{aligned}
[\delta \mid \cdot] &\propto \prod_{i=1}^{n} (\Lambda_i(y_i)\tilde{f}_i(y_i))^{\Delta_i} \exp(-\Lambda_i(y_i)(1 - \check{S}_i(y_i)))[\delta] \\
&\propto \exp \sum_{i=1}^{n} \left(\Delta_i \log\left(\Lambda_i(y_i)\right) - \Lambda_i(y_i)\left(1 - \check{S}_i(y_i)\right)\right)[\delta]
\end{aligned}$$

These densities do not have a standard form.

Acknowledgement. The authors deeply thank the editor, the associate editor and the referee for several suggestions and comments which have greatly improved the paper.

References

[1] J. Berkson, and R. P. Gage, Survival curve for cancer patients following treatment, *Journal of the American Statistical Association,* **47**(1952), 501–515.

[2] R. A. Betensky and D. A. Schoenfeld, Nonparametric estimation in a cure model with random cure times, *Biometrics* **57** (2001), no. 1, 282–286.

[3] E. R. Brown and J. G. Ibrahim, A Bayesian semiparametric joint hierarchical model for longitudinal and survival data, *Biometrics* **59** (2003), no. 2, 221–228.

[4] E. R. Brown and J. G. Ibrahim, Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine trials, *Biometrics* **59** (2003), no. 3, 686–693.

[5] B. P. Carlin and T. A. Louis, *Bayes and empirical Bayes methods for data analysis*, Chapman & Hall, London, 1996.

[6] M.-H. Chen, J. G. Ibrahim and D. Sinha, A new joint model for longitudinal and survival data with a cure fraction, *J. Multivariate Anal.* **91** (2004), no. 1, 18–34.
[7] M.-H. Chen, Q.-M. Shao and J. G. Ibrahim, *Monte Carlo methods in Bayesian computation*, Springer, New York, 2000.
[8] Y.-Y. Chi and J. G. Ibrahim, Joint models for multivariate longitudinal and multivariate survival data, *Biometrics* **62** (2006), no. 2, 432–445.
[9] B. J. Cowling, J. L. Hutton and J. E. H. Shaw, Joint modelling of event counts and survival times, *J. Roy. Statist. Soc. Ser. C* **55** (2006), no. 1, 31–39.
[10] M. K. Cowles and B. P. Carlin, Markov chain Monte Carlo convergence diagnostics: a comparative review, *J. Amer. Statist. Assoc.* **91** (1996), no. 434, 883–904.
[11] D. R. Cox, Regression models and life-tables, *J. Roy. Statist. Soc. Ser. B* **34** (1972), 187–220.
[12] U. G. Dafni, and A. A. Tsiatis, Evaluating surrogate markers of clinical outcome measured with error, *Biometrics* **54**(1998), 1445–1462.
[13] E. Demidenko, *Mixed models*, Wiley-Intersci., Hoboken, NJ, 2004.
[14] V. T. Farewell, Mixture models in survival analysis: are they worth the risk?, *Canad. J. Statist.* **14** (1986), no. 3, 257–262.
[15] C. L. Faucett, N. Schenker and J. M. G. Taylor, Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data, *Biometrics* **58** (2002), no. 1, 37–47.
[16] A. E. Gelfand and B. Mallick, Bayesian of Proportional Hazards Model Built From Monotone Function, *Biometrics*, **51**(1995), 843–852.
[17] W. R. Gilks, N. G. Best, and K. K. C. Tan, , Adaptive Rejection Metropolis Sampling within Gibbs Sampling. *Journal of Applied Statistics.* **44** (1995), no. 4, 455–472.
[18] A. I. Goldman, Survivorship Analysis when Cure is A Possibility: *A Monte Carlo Study, Statistics in Medicine,* **3**(1984), 153–163.
[19] R. J. Gray and A. A. Tsiatis, A linear Rank Test for Use when the Main Interest is in Differences in Cure Rates, *Biometrics,* **45**(1989), 899–904.
[20] R. Henderson, P. Diggle, and A. Dobson, Joint modeling of longitudinal measurements and event time data. *Biostatistics*, **4**(2000), 465–480.
[21] Hogan, J. W. and Laird, N.M. (1997). Mixture models for the joint distributions of repeated measures and event times. *Statistics in Medicine 16*, 239-257.
[22] J. G. Ibrahim, M.-H. Chen and D. Sinha, *Bayesian survival analysis*, Springer, New York, 2001.
[23] J. G. Ibrahim and P. W. Laud, A predictive approach to the analysis of designed experiments, *J. Amer. Statist. Assoc.* **89** (1994), no. 425, 309–319.
[24] J. G. Ibrahim, M.-H. Chen and D. Sinha, Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials, *Statist. Sinica* **14** (2004), no. 3, 863–883.
[25] D. Kalbfleisch and R. L. Prentice, *The statistical analysis of failure time data*, Second edition, Wiley-Intersci., Hoboken, NJ, 2002.
[26] Kuk, A. Y. C. and Chen, C. H.(1992), A mixture Model Combining Logistic Regression with Proportional Hazards Regression, *Biometrika 79*, 531-541.
[27] Law, N.J., Taylor, J.M.G., and Sander, H. (2002). The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics 3*, 547–563.
[28] H. Lin *et al.*, Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer, *J. Amer. Statist. Assoc.* **97** (2002), no. 457, 53–65.
[29] R. A. Maller and X. Zhou, *Survival analysis with long-term survivors*, Wiley, Chichester, 1996.
[30] Peng, Y. and Dear, K. B. G. (2000), A nonparametric Mixture Model for Cure Rate Estimation, *Biometrics 56,* 237–243.
[31] D. Sinha and D. K. Dey, Semiparametric Bayesian analysis of survival data, *J. Amer. Statist. Assoc.* **92** (1997), no. 439, 1195–1212.
[32] X. Song, M. Davidian and A. A. Tsiatis, A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data, *Biometrics* **58** (2002), no. 4, 742–753.

[33] D. J. Spiegelhalter *et al.*, Bayesian measures of model complexity and fit, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** (2002), no. 4, 583–639.
[34] J. P. Sy and J. M. G. Taylor, Estimation in a Cox proportional hazards cure model, *Biometrics* **56** (2000), no. 1, 227–236.
[35] Taylor, J. M. G. (1995), Semi-Parametric Estimation in Failure Time Mixture Model, *Biometrics 51*, 899-907.
[36] Taylor, J. M. G., Cumberland, W.G., and Sy, J.P. (1994). A stochastic model for analysis of longitudinal data. *Journal of the American Statistical Association 89,* 727–76.
[37] A. A. Tsiatis and M. Davidian, A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error, *Biometrika* **88** (2001), no. 2, 447–458.
[38] A. A. Tsiatis and M. Davidian, Joint modeling of longitudinal and time-to-event data: an overview, *Statist. Sinica* **14** (2004), no. 3, 809–834.
[39] A. D. Tsodikov, J. G. Ibrahim and A. Y. Yakovlev, Estimating cure rates from survival data: an alternative to two-component mixture models, *J. Amer. Statist. Assoc.* **98** (2003), no. 464, 1063–1078.
[40] Yakovlev, A. Y. and Tsodikov, A. D. (1996), Stochastic Models of Tumor Latency and Their Biostatstical Applications, Hackensack, New Jersey: *World Scientific.*
[41] G. Yin, Bayesian cure rate frailty models with application to a root canal therapy study, *Biometrics* **61** (2005), no. 2, 552–558.
[42] G. Yin and J. G. Ibrahim, A general class of Bayesian survival models with zero and nonzero cure fractions, *Biometrics* **61** (2005), no. 2, 403–412.